DOCUMENT RESUME

ED 196 430 IR 009 038

AUTHOR Cobb, Hugh
TITLE Data Base Management Needs of National Assessment and Ways to Meet Those Needs.
INSTITUTION Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO NAEP-12-IP-59
PUB DATE Jul 80
GRANT NIE-G-80-0003
NOTE 65p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Archives; *Computer Oriented Programs; *Data Analysis; *Data Bases; Educational Assessment; Information Dissemination; *Information Needs; *Management Information Systems; National Programs; Online Systems; Quality Control; User Satisfaction (Information)
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

This paper discusses the procedures best suited for storing, organizing, and using past and present National Assessment of Educational Progress (NAEP) data, as well as procedures and systems used for archiving past and future data, and the quality-control procedures employed to ensure the integrity and responsiveness of the data management system. Four categories of users of NAEP data are identified: external researchers interested in specific analyses of the data, external users interested in existing research results, archival users interested in preserving information gathered by NAEP, and internal users (the staff) with specific interests in part of NAEP data. A discussion of database management needs describes the needs characteristics of each of these categories of users and points out some of their differences and similarities. Database management techniques are explored in terms of equipment, software, methods, and applications of the NAEP data management technique. Extensive appendices include a list of the contents of a current NAEP public-use data tape, data access plan and dissemination plan for the NAEP, and an overview of data models. A bibliography is provided. (Author/MER)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED196430

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

DATA BASE MANAGEMENT NEEDS OF NATIONAL ASSESSMENT AND WAYS TO MEET THOSE NEEDS

No. 12-IP-58

by Hugh Cobb

Assisted by the Staff of the National Assessment of Educational Progress

Education Commission of the States
Suite 700, 1860 Lincoln Street
Denver, Colorado 80295

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

T. Pratt

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

July 1980

IR009035

The National Assessment of Educational Progress is funded by the National Institute of Education. It is under contract with the Education Commission of the States. It is the policy of the Education Commission of the States to take affirmative action to prevent discrimination in its policies, programs and employment practices.

The National Assessment of Educational Progress is an education research project mandated by Congress to collect and report data, over time, on the performance of young Americans in various learning areas. National Assessment makes available information on assessment procedures and materials to state and local education agencies and others.

The work upon which this publication is based was performed pursuant to Grant No. NIE-G-80-0003 of the National Institute of Education. It does not, however, necessarily reflect the views of that agency.

TABLE OF CONTENTS

PREFACE. 1

OVERVIEW 3
The Purpose of This Paper. 3
Basic Assumptions. 3
The Contents of This Paper 3

SECTION 1 DATA BASE MANAGEMENT NEEDS. 5
Introduction 5
Needs of External Researchers. 6
Needs of External Users of Existing Results 7
Needs of Archival Users of NAEP Data. 9
Needs of Internal Users. 10
Summary. 14

SECTION 2 DATA BASE MANAGEMENT TECHNIQUES 15
Introduction 15
Equipment. 15
Software 17
Methods. 19
Applications 22
Summary. 24

SECTION 3 CONCLUSIONS 25

APPENDIX A CONTENTS OF A CURRENT NAEP PUBLIC-USE DATA TAPE. 27

APPENDIX B DATA ACCESS PLAN FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS 29
Data Access Conference 29
Information Retrieval and Data Analysis. 31
Online Access. 32
Public-Use Data Tapes. 33
Data Archives. 36
Dissemination and Follow-Up Activities 37
Implementing the Data Access Plan. 38

APPENDIX C DISSEMINATION PLAN FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS 41
Audiences. 41
Range of Products. 41
Dissemination Network. 42
The Superintendent of Documents and the Government Printing Office 42
Library of Congress. 43
Educational Resources Information Center 43
Distribution Center. 44
Bimonthly Newsletter 45
Journals and Article Reprints. 45
Assessment Products. 45
Dissemination Strategies 46
Making More People Aware 47
Annual Assessment Conference 48

APPENDIX D PROCESS SUMMARY FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS DATA BASE MANAGEMENT. 49

APPENDIX E AN OVERVIEW OF DATA MODELS, WITH EMPHASIS ON THE RELATIONAL MODEL 53

APPENDIX F NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS RELEASE POLICY FOR CONFIDENTIAL DATA. 57

BIBLIOGRAPHY 59

LIST OF TABLES

TABLE 1. Classes of Information and Potential Users. 8

TABLE 2. Types of NAEP Data 11

TABLE 3. Internal Users of NAEP Data. 12

TABLE 4. Current NAEP Computer System Components 16

TABLE 5. NAEP Computer System Features and Utilization. 17

TABLE 6. Data Base Management Methods. 20

TABLE D-1. Creating and Loading Data Dictionaries 49

TABLE D-2. Defining Assessment Respondent Data File Structure. 49

TABLE D-3. Structuring and Loading Respondent Data Files. 50

TABLE D-4. Item-Level Data Analysis 50

TABLE D-5. Item-Level Trend Analysis. 50

TABLE D-6. Summary-Level Data Analysis/Reporting 51

TABLE D-7. Item-Level Data Reporting. 51

NOV 17 1980

PREFACE

The following paper is one of seven required by the National Institute of Education (NIE) grant to the Education Commission of the States for the administration of the National Assessment of Educational Progress (NAEP). The papers are intended to raise significant questions about NAEP functions and procedures in order to apprise the Assessment Policy Committee of its technical and administrative options as NAEP moves into a second decade of activity. According to the NIE Program Announcement, each paper should include "discussions about the significance of a national assessment of educational progress for (a) informing public opinion, (b) influencing educational and other social policies, (c) identifying specific educational and research needs and (d) contributing to the goal of improving education." Thus, the papers combine broad theoretical considerations with very concrete discussions of potential options and their trade offs. All the papers are suggestive, rather than definitive -- starting points for discussion rather than end points.

Two fundamental questions undergird these papers. First, if one were to design a national assessment program from scratch to serve the needs of the eighties and nineties, what would it look like? What theory would it draw upon? What constituents would it serve? How would it operate? And second, given that an assessment has been operating for 10 years, how do answers to these questions match the current NAEP? How could the current NAEP be changed to accommodate needs and technology unforseen when it was begun in the 1960s? As these questions weave in and out of the discussions of various issues, it will become clear that the answers one gives will probably depend upon one's assumptions about such things as the way society, education and science work, as well as one's familiarity with the theory, politics, history and operation of the current National Assessment.

There are many ways to implement the 1978 legislation mandating a national assessment. Some are undoubtedly better than others in certain respects; all have both advantages and drawbacks, complicating even the

simplest comparisons among alternatives. In such a situation, it is imperative that advocates for various models of a national assessment adopt an open, interdisciplinary approach to rethinking the Assessment, listening carefully to each other and respecting professional differences of opinion.

Individuals charged with making policy decisions about NAEP -- members of the Assessment Policy Committee and members of the U.S. Congress -- will find that, in many instances, attempts to strengthen the Assessment in one regard will weaken it in some other regard. Given a certain level of funding, advances in some areas come at the expense of others. The National Assessment can neither be all things to all people nor a single, unchanging enterprise. It will always represent a combination of various tensions. The question for its future is not how do we eliminate those tensions? Rather it is: How do we harness those tensions creatively to keep the Assessment abreast of the times and continually relevant to the needs of American education?

OVERVIEW

The Purpose of This Paper

This paper addresses three questions raised in the National Institute of Education (NIE) Program Announcement for the National Assessment of Educational Progress (NAEP): (1) "What data base management procedures are best suited for storing, organizing and using past and present data?" (2) "What procedures/systems may be used for archiving past and future NAEP data so that they are relatively easily available to the grantees, NIE and others for secondary analyses?" and (3) "What kinds of quality-control procedures will be employed to ensure the integrity and responsiveness of the data management system?"

Basic Assumptions

This paper uses the phrase "data base management system" in two ways. When capitalized, or abbreviated DBMS, it will refer to a software package (typically vendor supplied) that provides specific data access capabilities to the user. It is a tool used primarily by data processing specialists to implement application systems. When not capitalized, it will refer to the much larger set of procedures, methodologies, hardware and software used to manipulate information. It is in this latter context that this paper addresses the questions of the Program Announcement.

Readers of this paper are assumed to be knowledgeable both of NAEP in general and of the types of data NAEP collects in particular. They are not assumed to be experts in data base management, but are expected to have a general knowledge of the subject.

The Contents of This Paper

Before answering the questions posed in the NIE Program Announcement, one must first address three more basic questions: (1) What are the data under discus-

sion. (2) Who are the users of this data? and (3) To what uses are these users putting the data?

For the purposes of this paper, the answer to the first question is: "All of the documentation and data associated with the actual assessment." This can range from content objectives to performance data for item clusters. Specifically excluded are such types of data as internal fiscal data, scheduling data and dissemination data.

Given this answer to the first question, Section 1 of this paper addresses the second and third, describing users and uses of NAEP assessment data. Following this, Section 2 of this paper discusses general and specific data base management techniques that NAEP has used or should use to meet the needs discussed in Section 1, and thereby addresses the original questions for the Program Announcement.

There are several major areas relating to use of NAEP data that have been explored in papers previously developed by NAEP. Among these are (1) access to NAEP data, discussed in the 1979 Data Access Plan (Appendix B herein); (2) dissemination mechanisms for NAEP data, discussed in the 1980 Dissemination Plan (Appendix C herein); (3) enhancement of NAEP's utility in meeting the informational needs of NAEP's audiences, discussed in National Assessment's issue paper entitled Major Informational Needs ... (1980); and (4) policies regarding release of school, district and/or state identifying data to secondary users, discussed by the NAEP Assessment Policy Committee periodically. (Appendix F herein is the policy as of June 1980.) Inasmuch as these issues have been thoroughly explored in the above-mentioned documents, they are generally only alluded to herein.

SECTION 1

DATA BASE MANAGEMENT NEEDS

Introduction

What are the data base management needs for NAEP? This question depends totally on the question "Who are the users of NAEP data?" Uses of NAEP data are discussed in Major Informational Needs ... (1980). More specific discussions with researchers, information management specialists, project staff and the National Institute of Education (NIE) staff have resulted in the definition of four distinct classifications of users of NAEP data. These groups are:

1. External (to NAEP operations) researchers -- typically university-affiliated education researchers, statisticians, psychometricians or subject-matter specialists -- who are interested in specific analyses of the data collected by NAEP. Their emphasis is on new analyses of data as distinct from use of existing results.

2. External users of existing research results -- for example, federal agencies, national education or curriculum organizations, state and local assessment programs, legislators -- who are interested in the analysis results of the Assessment. These users may desire information at many different levels. For example, a state assessment may desire national comparison data on a group of items, whereas a legislator may need data on performance trends for population subgroups. The unifying attribute of this heterogeneous class of users is the need for existing NAEP information, as distinct from the development of new results from the NAEP data base.

3. Archival users of NAEP data who are primarily concerned with preserving the information gathered by NAEP for future users. Archiving for NAEP ranges from the storing of the machine-readable item response data (as a minimum) to

the warehousing of all original data collection instruments, project working memoranda, analysis specifications, analysis results and external correspondence. The archiving task is: (1) to determine the appropriate location on this continuum for NAEP and (2) to ensure that whatever is archived can be retrieved and used.

4. Internal users -- the NAEP staff -- comprise five subgroups, roughly parallel to the NAEP functional organization: development, analysis and research, publications and user products, user services, and data processing. The first four subgroups each have specific needs for parts of the NAEP data base. The fifth subgroup, data processing, is responsible for meeting these needs as well as the data base management needs of the other three major classes of users.

Each of these four classes of users has a unique set of needs that are discussed separately in the following sections.

Needs of External Researchers

What are the data base management needs of the external researcher? To help answer this question, NAEP sponsored a Data Access Conference in late 1979 that brought together experts in the fields of archiving, documentation and education research.[1] These experts, together with NAEP staff, delineated the following as the essential characteristics of external researchers:

1. They are experienced professional researchers, not students or novices.

2. Their goal in using NAEP data is to perform new analysis.

3. To implement this goal, they need machine-readable versions of the data collected by NAEP.

---

[1]This conference resulted in the development of a NAEP Data Access Plan, which is Appendix B of this paper.

4. They are knowledgeable and experienced in the use of standard statistical packages.

5. They have experience in performing complex data analysis.

6. They are not data processing or data base management experts.

7. They have a general knowledge of NAEP, but do not know specifics of the data content.

8. Their research is generally independent of ongoing NAEP work.

These characteristics specify to a large extent the external researcher's needs. For example, characteristics 3, 7 and 8 above lead to the conclusion that researchers need an independent machine-readable copy of NAEP data files (typically on magnetic tape). Furthermore, characteristics 4 and 6, together with the complex nature of NAEP data, dictate the use of rectangular data files that are organized by data collection unit and are easily accessible from standard statistical packages. Characteristics 7 and 8 imply that documentation of the data should be self-sufficient, and, to as great an extent as possible, machine-readable.

A collection of data files, with complete documentation, affords the researcher the necessary flexibility, while allowing NAEP to develop a general purpose "public-use data tape" schema applicable to all collected data. A schema along these lines was developed by NAEP staff during the first six months of 1980 and is the basis for NAEP's external researcher support today. Appendix A provides a detailed contents list for a prototype NAEP public-use data tape.

Needs of External Users of Existing Results

What are the data base management needs of the external user of existing research results? A very general answer to this question is, "Access to selected item- and summary-level analysis results and documentation, together with descriptions of organizing and classifying schemes for this information." Such access could in principle occur through a series of published documents or through interaction with the NAEP staff in the user services area, or through some "user-friendly" computer

system for retrieving information. (Note that these alternatives are not mutually exclusive. For example, NAEP service staff could make use of a computer-based information retrieval system to service external requests for information.)

What are the types of infomation that this group of users need to access? Table 1 depicts the classes of information present at NAEP, together with the potential users of the information.

TABLE 1. Classes of Information and Potential Users

Information	Users
Assessment objectives	Subject-matter specialists Curriculum specialists State assessments Local assessments
Item description, text, classifications and performance data	Subject-matter specialists State assessments Local assessments Curriculum specialists Classroom teachers
Summary-level performance data	Federal agencies National education organizations Legislators Media

Much of this information is currently available in NAEP publications. The full set of this information is currently available through the NAEP User Services Department.

Regardless of whether a user is obtaining information directly, or through the NAEP User Services or Public Information Departments, there is a need for a powerful information retrieval capability. The most significant attributes of the NAEP information resource are its depth and diversity. To access this information effectively, a computer-based indexing and classifying system is essential. Such a system, including documentation of the system's content, organization and use, is the nucleus of the information retrieval process.

What are the attributes of such an indexing and classifying system? First, it should be usable by a non-

computer-sophisticated individual. Second, it should be keyword oriented. (Naturally all extant keywords would be thoroughly defined in the ancillary system documentation.) Third, the iterative nature of most inquiries demands that it be interactive. Fourth, the system should be highly "user-friendly," i.e., tolerant of imprecision and helpful in error diagnosis and recovery.

What information does such an indexing and classifying system provide in response to a specific query? At the most basic level, such a system provides lists of relevant items, objectives, analyses, and so forth, together with a description of where each piece of information is located. If (as is likely to be the case for NAEP) such locations include references to computer files, the system can provide the necessary access specifications, and can even access the information for the user.

Needs of Archival Users of NAEP Data

What are the data base management needs of the archival user? Two relatively distinct subgroups compose the archival user group: (1) those users desiring to perform analyses using NAEP archival data and (2) those users interested in preserving NAEP data as a national resource.

Users in the first subgroup have data base management needs similar to those of both previous user groups with the additional requirements of (1) physical access to original data collection instruments and (2) access to NAEP methodological studies, working papers and technical memoranda. This access is often more a logistic than a data base management concern; consequently, a cataloging and indexing schema for source documents is a necessary feature of effective physical access.

A computer-based implementation of such a schema composes the major part of the second subgroup's data base management need. Expanding the cataloging and indexing data base mentioned above to include project memoranda, publications, analyses, data files and correspondence provides the framework for preserving NAEP information. A complementary need in this area is a schema for the physical storage of the information. Although this topic is beyond the scope of this paper, several issues in this area are worth mentioning:

1. What documents will be archived?

2. How will the various documents be archived? Possible choices include computer-readable files, microfiche, microfilm and hard copy.

3. What level of archive redundancy is appropriate?

4. What information about NAEP archives should be disseminated? How should it be disseminated?

There is a considerable overlap between the cataloging schema above and the indexing and classifying schema referenced in the section above discussing users of existing results. This connection is hardly surprising; in fact, the archivist bases the decision of what to archive on what information is sought by users of the data.

Needs of Internal Users

What are the data base management needs of the NAEP staff? As mentioned in the introduction to this section, there are five functional subgroups of the NAEP staff, each with its own unique data base management needs. The first four subgroups -- development, analysis and research, publications and user products, and user services -- are the internal users for the fifth subgroup, data processing. This section briefly discusses the functional needs of the first four groups, and then discusses in detail the underlying data base management functions necessary for the data processing group to service its users (both internal and external).

What are the functional needs of the internal user subgroups? The answer to this question consists of a description of the types of data with which NAEP staff work together with the ways in which NAEP staff access each data type. These types of data (shown in Table 2) consist of the types of data referenced by external users, plus a few additional types. It is hardly surprising that many of these types of data are used by more than one internal user subgroup within NAEP. Table 3 depicts the uses of the various types of data by the user subgroups.

TABLE 2. Types of NAEP Data

Objectives	Specification of subject-matter content areas
New assessment items	Exercises (with accompanying documentation) that have been developed but not used in an actual assessment
Existing assessment items	Exercises (with accompanying documentation) that have been used in one or more assessments
Tryout response data	Performance data for items (typically new assessment items) collected for a small sample of respondents for the purpose of evaluating the quality of the items
Analysis specifications	Detailed criteria for the analysis of assessment items. These include such information as correct responses, derived response specifications, item-cluster definitions, item scoring specifications and respondent demographic data code definitions
Respondent data	The item-by-item responses to assessment items. These are collected during an assessment; included in this category are background questions, demographic information, sampling data and principals' questionnaire responses
Item analysis results	Results of statistical analysis of the respondent data
Summary analysis results	Results of aggregation of item analysis results

TABLE 3. Internal Uses of NAEP Data

User Subgroup	Type of Data	Uses
Development	Objectives	Creation, update, retrieval
	New assessment items	Creation, update, retrieval
	Existing assessment items	Retrieval
	Tryout response data	Creation, analysis*
Analysis and research	Existing assessment items	Retrieval
	Analysis specifications	Creation, update, retrieval
	Respondent data	Retrieval, analysis*
	Item analysis results	Retrieval, analysis*
	Summary analysis results	Retrieval
Publication and user products	Item analysis results	Retrieval
	Summary analysis results	Retrieval
User services	Objectives	Retrieval
	Existing assessment items	Retrieval
	Analysis specifications	Retrieval
	Item analysis results	Retrieval

*The analysis functional use is the capability to use the data as input to a standard statistical package.

What are the functional DBMS needs of the data processing subgroup of NAEP? The data processing subgroup has two service roles: (1) to perform data processing and data management functions for NAEP and (2) to provide NAEP staff and the external users of NAEP with access mechanisms to the NAEP data bases. Both roles require a powerful Data Base Management System as a basis for efficient and effective processing of information. Such a system needs the following properties:

1. It must be able to organize and link diverse data bases. Since NAEP data occur at respondent, item and summary levels, and in fact recur periodically, the Data Base Management System must facilitate linking specific portions of the data files on request.

2. It must facilitate updating of the data. The dynamic, evolutionary nature of NAEP implies frequent modifications to data bases. Even respondent data are subject to change, in the course of rescoring open-ended items for trend analysis.

3. It must facilitate revision of file structures. When logical file structures need to be changed, a Data Base Management System serves two roles: (1) to implement the structure change easily and (2) to insulate user applications from the effects of the change, when possible (data independence).

4. It must facilitate report generation. Such capabilities as retrieval, sorting and standard report gereration should be integrated with the Data Base Management System.

5. There should be an ad hoc query capability for nonstandard or special reports. This feature is perhaps the most crucial in developing a system truly accessible to persons who are not data processing specialists.

6. Extensive, expandable data dictionaries should be integrated into the Data Base Management System, and should be accessible as data themselves. NAEP's need for thorough field-level documentation is paramount; the number of distinct fields and the need for level linkages (respondent to item to item cluster) necessitate being able to treat the field-level docu-

mentation as data.

7. It should provide dynamic host language access capabilities. NAEP performs both ad hoc and standard statistical calculations on the Assessment data. The complexity of these calculations dictates that they be programmed in a high-level procedural language. Ideally, these programs should access the field definitions on the data dictionary, and hence, the data themselves.

8. It should have extensive online query, update and host language processing capabilities. Such capabilities are at the heart of "user-friendly" effective access to the NAEP data.

9. There should be straightforward access to the data from standard statistical packages such as SPSS, SAS, BMD and OSIRIS. This capability facilitates ad hoc exploration of the data both by NAEP staff and external users.

10. It should foster application data independence. The ability to insulate existing applications from changes in data structure and access mechanisms is integral to NAEP's need for analysis flexibility.

The ways in which these needs are currently met and the ways in which NAEP proposes to increase its ability to meet these needs are the subjects of Section 2 of this paper.

Summary

This section of the paper has concerned itself with defining the various user groups of NAEP data and examining their respective needs. Although each group has specific, unique needs, there is a common element that pervades most of the users, namely, the need for a "meta" data base containing information about other data bases. Such a "meta" data base application is comparatively straightforward from a data processing perspective; however, the diversity and interdependenies in data base contents -- the actual information -- makes such a data base much more difficult to establish and maintain. Nonetheless, NAEP should consider expending effort in this direction.

SECTION 2

DATA BASE MANAGEMENT TECHNIQUES

Introduction

What data base management techniques are best suited to meeting the needs of the NAEP user community? This section of the paper will explore answers to this question in four areas: equipment, software, methods and applications. The first three of these are enabling mechanisms -- tools for efficiently and effectively implementing the fourth area, applications. Each area can be discussed abstractly, assuming no existing context for discussion, or concretely, assuming the existing NAEP data base management systems. What follows synthesizes these two approaches -- specifying the abstract capability, discussing to what extent the current NAEP systems provide the capability and discussing what steps are appropriate to accommodate the needs outlined in Section 1.

In most instances, there is a continuum of implementations that meet the needs of NAEP users. Not surprisingly, each such continuum has an associated cost continuum. Appropriate placement of NAEP on each continuum is the responsibility of the management of NAEP and, ultimately, the NAEP Assessment Policy Committee. As user needs and project priorities evolve, NAEP management must reevaluate NAEP's location on the various continua and allocate resources accordingly.

Equipment

What equipment does NAEP need to meet the data base management needs of its users? An abstract answer is "a medium-scale, general-purpose computer system, with local and remote online capabilities." Given the context of NAEP data processing as it has existed since NAEP's inception, a more concrete answer is "a medium-scale computer system compatible with IBM 370 architecture, with some variety of high-speed online local access, and some variety of lower-speed remote access."

The current NAEP computer system meets both definitions above. Table 4 describes the current hardware configuration at NAEP.

TABLE 4. Current NAEP Computer System Components (All Model Numbers Are IBM)

1	370 Model 138 Central Processor with 1 megabyte of storage
4	3330 disk drives (capacity 100 million bytes each)
8	2314 disk drives (capacity 29 million bytes each)
2	3420 tape drives (dual density 800-1600 BPI)
1	1403 printer (1,100 lines per minute)
1	2540 card reader/punch
13	3277 CRT terminals -- local
2	Dial-up communications lines (110 baud)
3	Control units (3272 for terminals, 3803 for tape drives and 2821 for printer and card reader/ punch)

Table 5 shows the degree to which the current system meets the NAEP user computer system needs (as presently defined).

TABLE 5. NAEP Computer System Features and Utilization

Features	Utilization Characterization
Processor speed	Currently adequate
Online disk capacity	Currently near maximum -- additional applications may necessitate expansion
Processor memory	Currently adequate -- additional online applications may necessitate expansion
Tape capacity	Currently adequate
Local terminals	Currently adequate -- additional online applications may necessitate expansion
Remote access capacity	Currently minimally adequate -- for all but nominal use, number of lines and speed of lines must both be increased

The current NAEP computer configuration is very well suited to most NAEP needs, with room for considerable expansion, as the need arises.

Software

What software does NAEP need to meet the data base management needs of its users? NAEP data base usage is a mixture of two extremes. At one end is the ad hoc user (internal or external) who has a unique analysis to perform. At the other extreme are the standard analyses generating data for primary analysis and reporting. The NAEP data base management system must effectively accommodate both of these extremes.

At the heart of NAEP's software needs is the need for a powerful Data Base Management System with extensive data structuring, storing, retrieving, querying and report writing capabilities. Ideally used for centralizing access to all NAEP data, such a system should include linkages to standard statistical software packages, interfaces to high-level languages and the capability of supporting online applications. Further, be-

cause a significant portion of both the external and the internal users are not data processing specialists, the Data Base Management System must provide nonprocedural access to the data.

Although the NAEP data base contains statistical survey data (ranging from respondent to summary level), the complexity and diversity of the data mandate the use of a programming language that is well suited to both numerical and nonnumerical applications. However, such a language addresses only a portion of the user interaction with the NAEP data. Widely used, nonprocedural statistical packages such as SPSS or SAS provide a complementary capability for ad hoc explorations of the data; NAEP data must be easily accessed via such packages.

Underlying all the software described above is the operating system for the computer system. NAEP data management needs require certain capabilities inherent in the operating system, namely, support of (1) online applications, (2) batch applications, (3) local and remote access and (4) a multiprogrammed environment.

The current NAEP software configuration has many, but not all, of the capabilities outlined above. The current Data Base Management System -- NIPS -- has effective structuring, storage, retrieval and report writing capabilities. However, it is based on a two-level hierarchical model of data. This model, while effective for many NAEP data bases, has proven limiting as the Assessment's data-structuring needs have become more complex. Further, neither the capabilities of linking data bases together, nor the interfaces to high-level languages (including statistical packages) are adequate for NAEP's present needs. These problems, together with the fact that NIPS is no longer a vendor maintained product indicate that NAEP should convert to a more powerful, vendor-supported Data Base Management System. Both because of the diversity of the NAEP data and because of the need for nonprocedural access to the NAEP data, a Data Base Management System based on the relational data model would be ideal. (Appendix E provides an overview of data models, with emphasis on the relational model.) Data Base Management Systems supporting a relational data model (or at least a relational view of data) are now commercially available; NAEP should evaluate alternatives in this area and install a new Data Base Management System as a prerequisite to fully meeting NAEP's user needs.

The programming language currently used at NAEP -- PL/1 -- is ideally suited to the diversity of NAEP applications. It is an excellent vehicle for top-down structured programming -- a must, given the open-ended nature of many of NAEP's applications.

NAEP's ad hoc, exploratory analysis needs are currently well addressed with SPSS, SAS, BMD and BMDP (all currently installed). These statistical packages provide a diverse set of data analysis capabilities, with largely nonprocedural specification languages. Currently, these statistical support packages interface to NAEP data through subset files extracted from the data base. Recently developed applications have afforded the capability of automatic file definitions for SPSS and OSIRIS. This capability can and should be extended to include SAS, as well as other widely used statistical packages.

NAEP currently uses the OS/VS1 operating system, with the Cullinane program product INTERACT. This combination has provided NAEP flexible, efficient capabilities for online program development, job submission, text editing, data entry, batch processing, and local and remote access, all in a multiprogrammed environment. At present, there is no way to implement true online applications -- to do so requires a data communications monitor (typically integrated with the Data Base Management System). Evaluation of potential monitors and acquisition of one suited to NAEP's needs is a necessary next step for NAEP to meet its users' needs.

Methods[1]

What data base management methods should NAEP use to meet its users' needs? Data base management methods are of two types: general and specific. General methods are techniques applicable independent of context, whereas specific methods are techniques that are appropriate for a particular application because of certain attributes of those applications. Table 6 details data base management methods of both types suitable for NAEP.

---

[1]Strictly speaking, use of a Data Base Management System, use of a high-level, block-structured programming language, and use of statistical packages are methods as well. Since these have already been discussed under software, they are not repeated here.

TABLE 6. Data Base Management Methods*

	Methodology	Type	Description
1.	Use of extract files for applications processing	Specific	Creation of subfiles off of a master data base for the purpose of achieving data independence and program simplification
2.	Use of generalized, parameter driven software	General	Development of applications programs that have few (if any) localizing parameters built into the program itself. Used to minimize the need for program redevelopment and to increase application flexibility
3.	Use of a data dictionary	General	Systematizing data documentation in program readable form to facilitate system modifications and enhancements
4.	Use of machine storage of analysis specifications	Specific	Extending the concept of data dictionaries to include statistical analysis specifications for the Assessment data.
5.	Use of automatic interfaces between extract files and statistical software packages	Specific	Creation of control statements describing input data files for SPSS, SAS, etc., for the purpose of relieving the user (especially the ad hoc user) of the (sometimes awesome) task of file definition

TABLE 6 -- Continued.

Methodology	Type	Description
6. Use of the relational data base model as an applications view of the data	General	Use of two-way tables as an underlying data structure in the application development process. Ideally used in conjunction with a relational data base management system

*What follows is a description of various data base management methods. Other methodologies, such as structured programming, top-down design and composite design, are vital to the success of NAEP, and are in use currently. They are not discussed here because they are not data base management related.

Although NAEP makes use of all the above in performing the data processing associated with the Assessment, the advent of a new (relational) Data Base Management System will substantially alter the ways in which NAEP uses the methods. For example, the data dictionary and machine storage of analysis specifications methods are currently implemented as an independent data base. A more powerful Data Base Management System will integrate most, if not all, of the capabilities inherent in these methods directly into the Data Base Management System. A more powerful way to present external views of the data (thereby insulating applications from the data complexity of the entire data base) will decrease the need to use extract files for processing. Finally, if NAEP obtains a relational Data Base Management System, NAEP will be able to fully integrate the relational data model into its applications processing. This is highly significant, since the nonprocedural data access mechanisms inherent in the relational model are an ideal vehicle for nondata processing experts.

Applications

What applications should NAEP have to meet the data base management needs of its users? The answer to this question is a direct consequence of Section 1 of this paper, which discusses users' needs. There are four major applications areas associated with the NAEP data: (1) standard analysis,[2] (2) special (or ad hoc) analysis, (3) external researcher support and (4) information retrieval (including textual data, analyses data, objectives, and so forth).

---

[2]Appendix D provides a brief process summary for the data management associated with standard analysis.

NAEP applications are meeting the project needs in the standard analysis area, primarily in batch mode. The recent acquisition of online capabilities, together with the planned acquisition of a more powerful Data Base Management System and a communications monitor, define the next appropriate enhancements in the standard analyses area. All of the NAEP data files (and related accessing programs) must be converted to operate under a new Data Base Management System. Further, the entire area should be reexamined to determine where online approaches should replace the batch approaches currently in use. The answer to this question is by no means clear-cut, since many of the processes involved are highly consumptive of processor time and do not require user interaction once initiated.

In contrast, ad hoc data analysis can benefit immediately from an online, interactive environment. The fundamental application needs in this area are the streamlining of the linkage between the NAEP data files and standard statistical packages, and the development of the capability for the data analyst to directly specify the subset of the NAEP data files that he/she desires to process.

The external research support area has undergone tremendous growth at NAEP in the past year. NAEP has developed a system for producing public-use data tapes that can be accessed independently. These tapes do an excellent job of meeting researchers' needs; however, there are areas that need improvement. Specifically, the public-use data tape system needs:

1. The facility to provide interface control card images for SAS.

2. More extensive text for item responses.

3. Methods for automatic conversion of scoring specifications from NAEP internal to external user formats (currently this is done manually).

4. An expanded cross-reference capability both of performance variables and of background variables, to further facilitate trend analysis.

5. A generalized, multisubject-area schema for classifying performance items.

The information retrieval area is the application area least supported by current NAEP systems. Presently, information retrieval needs are met by allocating staff resources (either data processing or user services, as appropriate) to specific information retrieval tasks. This limits the user support that can be provided.

The indexing and classifying system described in Section 1 in the subsections entitled "Needs of External Users of Existing Results" and "Needs of Archival Users of NAEP Data" is the nucleus of a NAEP information retrieval system. Applications involving retrieval of specific pieces of information may be incorporated into such a system, on a priority basis. For example, access to item classification information may be deemed to be most important, and therefore might be implemented first.

Summary

This section of the paper has dealt with specific ways in which NAEP meets user data base management needs, and has discussed the ways in which NAEP data management should evolve to better meet the users' needs.

For the most part, NAEP is meeting its current users' needs. However, there are several specific areas that NAEP must address. It should evaluate available vendor supplied Data Base Management Systems and acquire one well suited to its needs. A similar (and parallel) evaluation and acquisition of data communications monitors is necessary for NAEP to develop effective online applications. Having acquired both of these software enhancements, NAEP should consider developing a basic indexing and classifying application, to be followed by selective information retrieval application.

SECTION 3

CONCLUSIONS

NAEP data base management supports a highly diverse group of users, ranging from the legislator who wants education achievement information to help define policy, to the researcher who is interested in discerning performance trends in mathematics. The data base management needs (especially in terms of applications systems) of these users are just as diverse as the group of users, ranging from online queries to interfaces with statistical packages.

The NAEP staff, and in particular the NAEP data processing staff, is responsible for meeting the data base management needs of NAEP's users. Although the data base management techniques currently in use serve the NAEP user community well in most respects, there are several steps that NAEP should take to complement existing capabilities:

1. Transition to a more powerful Data Base Management System (DBMS), since staff time invested in applications development prior to acquisition of a new DBMS will add to the transition burden

2. Acquisition of a general communications monitor

3. Enhancement of linkages to statistical packages

4. Development of an interactive indexing and classifying system (ideally occurring in the context of a new DBMS)

NAEP staff will be recommending all of these steps to the Assessment Policy Committee, subject to budget constraints.

The needs of the NAEP user community are highly dynamic; NAEP data base management must be able to accommodate these evolving needs. Equally dynamic are the priorities of service to various user communities, as established by the Assessment Policy Committee. Using

the flexible tools and methods described herein, the NAEP staff will continue to respond to the changing needs and priorities.

APPENDIX A

CONTENTS OF A CURRENT NAEP PUBLIC-USE DATA TAPE

A NAEP public-use data tape contains the machine-readable data collected during an assessment for a specific age group of respondents. Each tape is accompanied by a printed copy of a User Guide for the tape and a set of attachments:

1. An objectives booklet for the assessment learning area(s)
2. A procedural handbook and/or technical report
3. Copies of exercise booklets, scoring guides and principal's questionnaires on microfiche
4. A NAEP Publications List
5. A tape table of contents list
6. A listing of the entire User's Guide and of the first few records of each other file on the tape
7. An errata sheet

The data tape itself contains the following information:

1. A print-image file containing the User Guide for the tape (identical to the hard copy provided to the user)
2. A print-image set of User Guide Appendixes containing
   a. A glossary of NAEP terminology
   b. Supplemental definitions of variables
   c. Item classification schema definitions
   d. Item classifications (for the schemas in c)

e. A cross-reference listing of item occurrences (across assessment ages and years)

f. Record layouts for the SPSS, OSIRIS and Data Interchange File card image files included on the tape

g. A description of sampling weight adjustment procedures used by NAEP

h. A description of complications in statistical analysis of NAEP data due to the sample design, together with methods of dealing with these complications

i. An explanation of how NAEP estimates standard errors

3. A print-image data dictionary (a highly abbreviated codebook) for each exercise booklet

4. A print-image codebook for each exercise booklet documenting of all variables contained on the respondent data file for the booklet

5. SPSS control card image files describing the variables contained on the respondent-level data file for each exercise booklet[1]

6. OSIRIS-IV card image files describing the variables contained in the respondent-level data file for each exercise booklet

7. A data interchange card image file of variable descriptor records that describe the variables contained on the respondent-level data file for all exercise booklets

8. A separate, rectangular, respondent-level data file for each exercise booklet in the assessment

---

[1]A planned future enhancement is the inclusion of SAS control card image files describing the variables contained on the respondent-level data file for each exercise booklet.

APPENDIX B

DATA ACCESS PLAN FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

(Note: This appendix consists of a data access plan developed by NAEP and submitted to NIE in December 1979).

After more than a decade of assessing the knowledge and attitudes of American students, the National Assessment of Educational Progress has evolved to a point at which its data are valuable to scores of users. It is NAEP's intention to make its data widely available and to assist in its use, but the project has had to answer a number of questions in order to achieve this. For instance: What kinds of users will want access to the data? What kinds of data will they wish? What services do they require from NAEP and how will NAEP provide those services? What specific technical problems will NAEP encounter in trying to establish those services, and what mechanisms should NAEP institute to address those problems? And, finally, how will NAEP know that it is providing the services acceptably?

Data Access Conference

To help answer these questions, NAEP sponsored, on November 29 and 30, 1979, a Data Access Conference that brought together experts in the fields of archives, documentation and data provision to secondary users; researchers currently using NAEP data; and selected NAEP staff. The purpose of the conference was to identify both what users need and how their needs can be met. The conferees and NAEP staff identified possible uses of NAEP data, alternate ways of providing it and problems that NAEP has encountered or may encounter in trying to provide that data. Specifically, the conferees and staff agreed upon the following major points:

Types of users. NAEP serves two major classes of users: those who wish to retrieve informational data or data from previous analyses, and those who wish to access respondent-level data to perform new analyses.

Access mechanisms. There are three major types of access mechanisms that NAEP could provide. They are: online access mechanisms for users interested in retrieving data already analyzed, public-use data tapes for those interested in performing new analyses and archival data (not necessarily machine-readable) for researchers who need access to original NAEP materials.

Feedback mechanisms. Regardless of the specific access mechanisms NAEP selects, they must include systematic ways for users to communicate their problems to NAEP and systematic ways for NAEP to update the information users have.

Dynamic system development. In adopting a data access system, NAEP should not be limited by the hardware currently available. All systems ought to be designed so that they are dynamic and can be easily adjusted to accommodate changes in hardware environment or user needs.

Usability of the data. NAEP data can provide valuable insights into the state of American education, but the data can be misused. It is National Assessment's responsibility to provide sufficient information to minimize the chance that NAEP data will be misused or will lead to invalid conclusions.

Professional expertise. As it implements an access plan, NAEP should use consulting experts in archives, information retrieval, online access and documentation.

Promotional plan. NAEP should search out potential users and let them know what is available, how they may access it and where they can obtain the training necessary to use it. The conferees suggested that NAEP investigate alternative dissemination strategies that make use of external organizations and products.

Confidentiality of data. The mechanism specified in special provision 1 of the NAEP grant is appropriate for resolving the issue of confidentiality of data.

It is upon these major findings of the conference, other points brought up by the conferees and its own knowledge and expertise that NAEP based its plan for increasing the accessibility of both its analyzed data and its raw data.

Information Retrieval and Data Analysis

There are three types of users of NAEP data (or any other kind of data, for that matter). On the one hand, there are those users -- policy researchers, for example -- who need to access specific items of information at a summary level. In the lexicon of NAEP, such information would be at the level of "Do females perform better than males in mathematics and is their performance related to courses taken?" or the like. On the other hand, there are those users -- education data analysts, for example -- who need NAEP raw data to do their own analyses because they ask questions that existing NAEP analyses simply do not answer.

The line between these two types of users, though conceptually clear, is often somewhat blurred in reality. Education data analysts often begin by asking questions that can be answered by simply retrieving available information, but their questions become more and more difficult to answer in this way. That is, at some point, they shift from information retrieval to data analysis. And depending on the mode in which the analyst is operating at any given time, there are different needs. Data analysis requires a totally different set of tools and access mechanisms than simple information retrieval. Where retrieving predetermined pieces of information requires a set of direct-access (or "online") mechanisms for accessing predetermined sets of data, data analysis requires raw data and supporting documentation that allow users to access the data on the computer hardware available to them.

This dichotomy between information retrieval and data analysis -- and the more subtle relationship between the two -- forms the base of NAEP's Data Access Plan, which addresses both areas. For information retrieval, the Plan requires investigating and adopting online mechanisms. For data analysis, there are provisions for developing public-use data tapes with full supporting documentation.

In addition, there is a third type of user with still different needs. This type of user -- for example, an historical or methodological researcher -- requires access to original documents (such as original student writings or artwork), rather than summary-level data. NAEP's plan addresses this need also by offering a comprehensive mechanism for providing archival data.

Online Access

For the purposes of the Data Access Plan, "online" refers to any direct access to NAEP data, whether or not the specific material is machine-readable. The attendees at the Data Access Conference identified a continuum of approaches that they would accept as reasonable, ranging from a published "telephone book" of NAEP data with extensive cross-indexing as the minimally acceptable device, to the ideal of a state-of-the-art, online computer system capable of handling several dozens of users simultaneously. Points along this continuum would provide acceptable levels of service for such users as school officials, congressional staff, agency staff or media; for initial access by researchers; and for internal exploration of data. Two of the conferees -- Richard C. Roistacher of the Bureau of Social Science Research and Sheldon Laube of Leinwand Associates -- suggested an online computer system that, in essence, maintains cross-referenced lists on computer files and allows simple access to them via a text editing system.

There is a similar continuum for the information the online system could provide. These contents include, but are not limited to, the following items:

- Objectives
- Exercise descriptions, texts and classifications
- Years, ages and package information for each exercise
- Summary analysis information such as p-values, national results, regional results, and so on
- Links between all levels of information
- Subject-area cross-references
- References to published documents
- ERIC accession information

Implementing an Online System

NAEP could choose from a wide range of online access systems. The conferees and staff, however, identified several problems that should be resolved before any such system could be implemented. Those problems are:

- Identifying the kinds of information to provide
- Selecting an information retrieval system
- Finding ways to avoid getting "locked in" to specific software
- Identifying the problems of, and acquiring the expertise for, developing natural language access systems
- Keeping users informed of problems in the data

It was further suggested that before implementing an online access system, NAEP should:

- Investigate what others have done to provide online access to data
- Hire consultants to evaluate NAEP's information retrieval needs
- Consider using an outside commercial information vendor to provide public access to the NAEP data base
- Expand the use of library search services, such as ERIC

Some conferees suggested that NAEP may not want to provide online access for external users itself, but might prefer to use an outside commercial information vendor to provide public online access. The cost of such a service bureau, Alice Robbin and others pointed out, would be comparatively modest once the data are documented.

Public-Use Data Tapes

Researchers who need raw data to perform their own analyses are the users most likely to want public-use data tapes containing respondent data. Such researchers will be neither students nor novices. They will have Ph.D.s, will be knowledgeable and experienced in the use of standard statistical packages like SAS and SPSS, and will have had experience performing complex data analysis. However, experience with complicated analyses is not equivalent to experience with NAEP specifically. This lack of specific NAEP experience, coupled with the

probability of experience with other complex data, has implications for the level of documentation NAEP should provide with its data tapes. (Thus, for example, the documentation must explain NAEP thoroughly but need not try to explain common statistical methods.)

The conferees identified a range of acceptable levels of documentation. The lower end of that range is a basic "data diet" -- just enough documentation (e.g., basic variable and category information) to allow an experienced researcher to use the file. The upper end of the range involves providing more extensive variable information, full category information, summaries of existing analyses, and more -- enough information to make the documentation totally self-sufficient. There are some technical problems involved with both ends of the range. At the lower end, the documentation could be developed comparatively quickly, but would necessitate a comprehensive training plan and the extensive availability of NAEP staff to researchers seeking answers to technical questions. The upper end of the range would decrease dramatically the necessity for researchers to rely on the availability of NAEP staff, but would greatly increase the time and resources needed to develop documentation. In fact, the conferees suggested the upper limit of the range may not be practically achievable.

The conferees agreed that the feedback mechanism is one of the most important elements of the documentation plan. Any first attempt at documentation is bound to be inadequate in some ways; the only way to improve it is to make it easy for users to comment on both the form and content of the documentation, and to make sure that their comments fuel a systematic recycling mechanism for updating existing documentation and issuing new editions of documentation.

Another type of public-use data tape would contain exercise analysis data, rather than respondent data. This kind of tape would most likely be used by curriculum specialists, policy analysts and subject-matter researchers. It, too, would require documentation, though of a different kind. As with the respondent data tape, feedback about the utility and efficiency of the documentation would be most important.

Implementing Public-Use Data Tapes

Although they identified a range of possible formats for data tapes, the conferees endorsed a basic plan for the format of NAEP's data files and accompanying documentation. They pointed out, however, that this plan was a starting point only, and that some areas of the plan would need to be reevaluated and adjusted as NAEP gains experience with its user community.

Some problems remain to be resolved, but resolution can occur only after interaction with the users of these initial products. The conferees helped NAEP identify problem areas, among which were:

The usability of the files. NAEP must make the use of its data tapes as easy as possible. To do this, NAEP must identify problems encountered in trying to use the initial products. Among the questions to be answered are these: Is the documentation clear? What changes to the product are necessary or desirable? Where should NAEP concentrate its efforts to upgrade its products?

Document processing system compatibilities. NAEP must investigate how much work will be involved in making the machine-readable documentation compatible with various commercially available document processing systems.

Standards. NAEP must decide on the standards it will follow in producing and distributing user tapes. How many types of tapes should NAEP distribute? To what degree will NAEP be able to honor special requests? Will NAEP be able to do custom analyses?

Trend data. NAEP must decide on the best way to describe how comparability of data was maintained across assessments so that secondary users can perform trend analyses.

Training. NAEP must find ways to train secondary users in unique aspects of NAEP data, or unique ways in which NAEP utilizes its data.

Distribution of user tapes. NAEP must evaluate whether it is the proper agent for distributing its user tapes. If it is not, who is and what level of tapes should NAEP provide them?

Data Archives

The conferees pointed out that some researchers may need access to original NAEP documents -- original data collection instruments or assessment packages, for example -- for historical or methodological research. They also stressed, however, that not all the original documents would be of interest to anyone and that NAEP must sift through its present archives. They suggested further that NAEP utilize consultants or archival experts to help decide what to do with its present archival data.

NAEP currently stores archival documents both on-site and at contractor facilities. These documents include:

At Westinghouse DataScore Systems:

- Essays
- Short answers
- Math work sheets
- Music tapes
- Art drawings
- Work flow and inventory control documents
- Editing/processing files/documents

At Research Triangle Institute:

- Administrative data collection forms
- Administrative procedures manuals
- Sampling frame data
- Administrative quality check records

At the Education Commission of the States:

- Development archives
- Analysis and reporting archives
- Contractor reports
- Dissemination archives
- Data processing archives
- Administrative archives

The conferees suggested that NAEP first compile a list of all archival documents in its and its contractors' possession, and then hire archival experts to help review the list, evaluate the documents and decide how to proceed.

There are several problems to be solved and many questions to be answered before NAEP can implement its

data archives. NAEP must decide what to make available routinely, what to keep for archives, what paper files to organize and maintain and what to document. Beyond that, there is the problem of how to maintain the archives. Should NAEP keep simple hard copies of the documents, or should it move to some other storage method such as microform? But perhaps the most perplexing problem of all is the problem of cross-references. NAEP's archives will not be useful unless the contents are adequately arranged, indexed and cataloged, and a catalog made available to all potential users of the archives.

Some conferees warned that any discarding of paper files will cause problems for future users. All decisions in archiving should be made with this in mind.

Dissemination and Follow-Up Activities

Dissemination is a vital part of the access plan. National Assessment must tell possible users that the data exist before a community of users can be built. NAEP must devise means to monitor the number and types of references to its data made by this community. And finally, NAEP must identify the problems users encounter with its data, so that access mechanisms may be improved.

The conferees suggested that NAEP follow a dissemination approach that (1) as an initial measure to build a user community, promotes the use of NAEP data to narrowly targeted audiences and (2) broadly disseminates information about the availability of NAEP data through existing organizational structures and outlets. Among the mechanisms the conferees suggested are the following:

- Develop brochures that organizations such as the Inter-University Consortium for Political and Social Research might distribute to universities and targeted audiences. These brochures would announce the existence of available and future materials

- Participate in meetings of groups whose members might have a use for NAEP data

- Conduct training session at both ad hoc and national or regional conferences

- Publish announcements in professional journals and publications

- Continue to use the ERIC system as fully as possible

Whatever mechanism or combination of mechanisms NAEP chooses, it must include a system for obtaining user feedback and utilizing this feedback to improve services. To do this, of course, NAEP must be able to identify specific users so it can approach them for suggestions and criticisms. The conferees suggested the following methods of identifying users:

- Encourage users to fill out user request, feedback, comment and evaluation forms so NAEP can distribute new or updated documentation

- Request that publications involving uses of NAEP data carry citation and acknowledgment statements

- Use organizations such as the Institute for Scientific Research, that can perform computer scans for authors and titles of papers by selected keywords, to monitor possible uses of NAEP data

- Conduct follow-up evaluations on the uses identified by the foregoing

Implementing the Data Access Plan

Implementing the three approaches detailed above -- online access, public-use data tapes and data archives -- will not be a simple task, due in part to NAEP's inexperience at providing secondary user data and in part to the scope of the work. Still, it is clear that NAEP's responsibility is to begin to build a community of users and to serve those users in an efficient manner. The experts at the Data Access Conference suggested that the most efficient approach to expediting this would be to:

- Emphasize the development and production of public-use data tapes

- Explore options for providing online public access to NAEP data

- Investigate the problems, needs and possible solutions to providing accessible data archives

Given the scope of the work necessary to implement this plan, and given its present level of experience and expertise, NAEP agrees with these conclusions. NAEP will utilize its present expertise in the area in which there is most likely to be the greatest demand -- public-use data tapes -- while enlisting the aid of outside experts to help investigate and weigh the options for various online access and data archives systems.

APPENDIX C

DISSEMINATION PLAN FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

(Note: This plan is for the calendar year 1980. Reference to specific dates herein are indicative of dissemination activities in general. This appendix consists of a dissemination plan developed by NAEP and submitted to NIE in March 1980.)

The National Assessment of Educational Progress (NAEP) generates a large volume of information. To make this information as useful as possible, National Assessment disseminates a variety of products designed to meet the needs of several different audiences.

Audiences

National Assessment has identified a number of major users of Assessment information: education administrators and curriculum specialists, education researchers and persons interested in replicating Assessment items and procedures for comparison purposes, the U.S. Congress and education policy makers, educators in a subject-matter area -- from local elementary and secondary teachers to those involved in teacher preparation -- professional associations, concerned lay persons and the general public.

Range of Products

The primary product of each assessment is performance data on how well young people are performing and how their performance has changed over time. These data are presented in several types of reports. Selective, or topical, reports summarize findings in lay language and in graphs appropriate for media representatives, legislators and the general public. Overviews -- for the same audiences -- give a broad sweep of the overall results of the assessments. Technical reports, displaying item-by-item and summary data, are available for

education researchers, statisticians or those interested in replicating Assessment items. National Assessment also makes available the objectives used to structure each assessment and the items (about half) used in a particular assessment. (The remainder of the items are kept secure so they can be used again to measure changes in performance over time.) National and group data for the released items are available for those interested in item-by-item results or replication. Data for secondary analyses by education researchers are provided on public-use data tapes (computer tapes with documentation). Procedural handbooks, describing procedures used in conducting a particular assessment, are available for researchers, those persons especially interested in methodology and those wishing to replicate NAEP items.

Dissemination Network

Through its 10-year history, National Assessment has developed and refined a dozen basic components of a dissemination plan. This plan, or network, includes close working relationships with the Superintendent of Documents; the Library of Congress and selected libraries around the United States; the Educational Resources Information Center as well as its own public information program; a bimonthly newsletter going to more than 34,000 people; an in-house distribution center; listings with Congressional Information Service, the NIE-funded Research and Development Exchange, and the Educators Progress Service, Inc.; publications; and various other directories. Recently, NAEP has become a part of the International Standard Book Number (ISBN) system, and every publication printed or reprinted by National Assessment will carry the widely used ISBN. The dissemination network is extended considerably by National Assessment's parent body, the 350-member Education Commission of the States, and the persons attending the annual large-scale assessment conference.

The Superintendent of Documents and the Government Printing Office

National Assessment publications are procured and printed through the Government Printing Office (GPO). The Superintendent of Documents (SOD) at present adds 790 copies to NAEP print orders plus any sales copies deemed advisable. About 670 of the 790 copies are sent to depository libraries around the country and abroad. (See next section on library distribution.) The SOD

selects and stocks about 20 percent of the NAEP results reports. NAEP reports are regularly included in the Monthly Catalog and every so often in Selected U.S. Government Publications. An SOD number is assigned to reports (such as HE 19:312) sent to depository libraries, and prices are set for NAEP reports. On request, the SOD furnishes sales and inventory reports to NAEP.

Library of Congress

Twelve copies of each NAEP publication printed at the GPO go to the Library of Congress (LOC). Two go into the permanent collection, one to Congressional Research Service, two to the duplicating section and the rest to the Exchange and Gifts Division. Objectives booklets are entered under series number LB 1570.N24, although the earliest ones were assigned individual numbers. Results reports are entered under series number LA 217.N364, although, again, early reports received separate numbers. Most depository libraries pick up the SOD number (HE) rather than the LOC number, although any library is free to use the Dewey Decimal, OCLC, MARC or their own numbering system. NAEP reports are available to congresspersons, staff and the public through the LOC. Through its Documents Expediting Project, the LOC distributes 130 NAEP Newsletters to selected libraries around the country.

The 670 copies sent to depository libraries include approximately 600 to public, university and state libraries; 46 to libraries abroad; 2 to the U.S. Senate Library; 2 to the U.S. House of Representatives Library; and 3 to the National Archives. This broad distribution through various kinds of libraries is an important and basic component of NAEP's dissemination network and plan.

Educational Resources Information Center

The current NAEP Publications List witnesses the close relationship NAEP has had with the Educational Resources Information Center (ERIC) over the years. All but a very few NAEP reports have been selected by ERIC, and all objectives booklets, monographs and many papers are now listed in Resources in Education and have ED numbers. In recent years, NAEP has produced duplicated exercise sets, sometimes with computer printouts reduced to 8"-X-10" size. These sets contain the items, documentation, scoring guides and national data. ERIC has

entered one of these into its system and others are pending or are in process. The NAEP Publications List displays the ERIC ED number with each product.

ERIC's indexing and abstracting procedure -- in addition to making all NAEP publications available in microfiche or hard-copy form -- is a valuable part of the dissemination plan. The fact that 16 clearinghouses and 700+ depositories worldwide have the ERIC collection again extends the NAEP dissemination network substantially.

Distribution Center

All NAEP products are available through the NAEP Distribution Center. Products available through the Distribution Center are listed in the NAEP Publications List, at present, citing more than 230 products. The publications list is periodically revised and reprinted to reflect the availability of new products. Products from previous assessments are reprinted according to demand so that data and information about previous assessments remain available. The NAEP Distribution Center has been in operation since November 1977. Prior to that time, NAEP publications were sent out without charge with no means of recouping production costs. During 1979, the Center received more than 7,900 requests, which resulted in sending out over 52,000 different items last year (excluding 203,875 copies of the NAEP Newsletter). Approximately 36 percent of the requests were for products involving a small cost; the remainder were items or materials that are free of charge. The bulk of the requests are from state and local education personnel, followed by requests from individuals and requests from national agencies. Leaflets and brochures are the items most often ordered, followed by subject-area summary reports and back issues of the newsletter. An inventory of products from previous assessment years exists; therefore, attempts are being made to reduce this inventory by promoting selected materials through the newsletter and other appropriate means. For example, reports from the 1972-73 mathematics assessment and a complete set of first assessment objectives booklets are being offered at a discount during 1980.

Bimonthly Newsletter

The NAEP Newsletter, a four- to six-page newsletter, has a circulation of more than 34,000. Recipients include legislators and education policy makers, state and local educators, education agency personnel, members of the media and individuals within the general public with an interest in education. People on the NAEP Newsletter mailing list are added in a variety of ways: NAEP staff and assessment administration field staff suggest people who would be interested in receiving the newsletter; those requesting materials are often asked if they would like to receive the newsletter; and consultants in various subject areas are regularly added to the mailing list, as are legislators at the state and local level and selected members of the media. The newsletter describes recent Assessment findings and products, uses of NAEP materials and data by secondary users and news about National Assessment. A recent purge of the newsletter mailing list was carried out, resulting in a slightly smaller but "cleaner" list.

Journals and Article Reprints

As schedules and funding allow, staff are encouraged to write journal articles concerning phases of the Assessment with which they have been particularly involved. Articles about NAEP are reprinted for dissemination. Staff works with national education agencies, such as the National Council of Teachers of Mathematics (NCTM) or the Association for the Supervision of Curriculum Development, making data readily available to them for their interpretation or dissemination. NCTM, for example, already has mapped out a series of interpretive articles on the second assessment of mathematics (1977-78) that will appear in Arithmetic Teacher, Mathematics Teacher, Phi Delta Kappan, Instructor and possibly in the Journal for Research in Mathematics Education during 1980.

Assessment Products

Upon completion, one copy of each Assessment product will be sent to the National Institute of Education for placement in the NIE library, and two copies will be submitted to the Educational Resources Information Center. An abstract describing the publication will be submitted with the product. In the case of nonprinted material, such as the public-use data tapes, a descrip-

tive abstract will be provided for placement in the NIE library.

Printed materials will continue to be produced by the Government Printing Office. Some items, for which small production runs are anticipated, may be duplicated by the Education Commission of the States' print shop. Computer-produced items, such as the public-use data tapes, will be generated by NAEP's Data Processing Department.

Dissemination Strategies

The writing change report, which is the first major summary report concerned with the third assessment of writing, will be released at a news conference. The report also will be announced and highlighted in the NAEP Newsletter, and a news release and a copy of the report will be mailed to selected members of the media. Approximately 300 members of the media will receive copies. Reports will also be mailed to federal and education agencies concerned with writing performance.

Following the release of the writing data, it is anticipated that NAEP staff will be requested to present findings to congressional committees. At times, congressional hearings are held to receive data from current research, and National Assessment is invited to testify at these hearings. Findings are discussed as well as ways in which professional associations are planning to make use of the information. National Assessment representatives have also been called on to testify at other hearings, e.g., when Congress was considering the reauthorization of the ESEA, NAEP representatives were heard by both House and Senate subcommittees.

The major dissemination strategy for NAEP publications will be to announce their availability in the NAEP Newsletter and, when appropriate, to run articles in the newsletter summarizing their content or their uses. Products will be available, at printing cost, through the NAEP Distribution Center, located in the National Assessment offices.

To help people determine the potential utility of products to them before ordering, NAEP makes available a series of leaflets, available at no cost, describing summary reports for a particular assessment.

A continuing part of the dissemination strategy is keeping participating schools informed. Each year, copies of the appropriate reports and pertinent information are sent to superintendents and principals of schools that have taken part in an assessment. Copies also go to consultants who have been involved in the development of objectives or exercises in the respective learning areas.

National Assessment is often represented at national conventions through personal presentations or printed materials. In 1980 it is anticipated that NAEP staff will participate in several national education meetings, such as the National Council of Teachers of English, the National Council of Teachers of Mathematics, the American Education Research Association and the Association for the Supervision of Curriculum Development.

Making More People Aware

Through its Public Information Department, NAEP has developed a close working relationship with the media, both print and electronic. Each time a new report comes out displaying new data about an assessment, the Public Information Department sends out a news release and a copy of the report to from 50 to 500 media representatives, depending on the learning area. The recent math findings, for example, were reported by more than 500 newspapers in the United States, reaching a total constituency of more than 50 million persons. Coverage by radio and TV is also increasing.

Although National Assessment results and methodologies are becoming known to an ever-increasing number of people, there are still all kinds of individuals -- working within the education community or not -- who could benefit from knowing more about the achievement of young people in subjects normally taught in the schools. An effort to make the findings better known is underway and will be expanded in 1980. Mailing lists are being solicited from education associations, teacher groups and commercial sources so that announcements of new NAEP materials -- tailor-made to the individual -- can be made to them directly. Others, as mentioned earlier, will receive notice of reduced rates for some of the earlier Assessment materials.

It is hoped that NIE will work in harness with NAEP in disseminating information about the Assessment.

Through its clearinghouses, laboratories, related projects, and its Testing, Assessment and Evaluation units, the National Institute of Education can play a valuable complementary role in the dissemination of National Assessment information.

The availability of materials for education researchers may also be publicized through appropriate channels. Brochures related to research materials will be available from NAEP. NIE might publicize the availability of this material; it might also be announced or publicized through AERA.

Annual Assessment Conference

National Assessment annually conducts a Large-Scale Assessment Conference in June. Persons involved with large-scale assessments at the national, state and local levels are invited to attend. The conference provides an opportunity for participants to exchange views, to listen to speakers on current issues of assessment and evaluation, and to benefit from each other's experiences. Materials related to the conference are mailed to approximately 1,200 potential attendees prior to the conference. Conference activities are publicized in the NAEP Newsletter. The approximately 300 who attend learn a great deal about National Assessment, which in turn extends the impact of NAEP throughout the states.

For the 1980 conference, an additional day was added to inform attendees of the newest developments in assessing writing skills. Sponsored jointly by NAEP and the NIE-sponsored Clearninghouse for Performance Testing at the Northwest Regional Educational Laboratory, the session focused on practical ways states and schools can use NAEP's methods to assess students' writing skills.

A variety of other dissemination activities will be carried out in 1980, such as NAEP participation in the National Science Foundation workshops (dissemination to researchers), Federal Register announcement of open Assessment Policy Committee meetings, announcements in Education Commission publications, and so forth. As needs become apparent, NAEP will respond.

APPENDIX D

PROCESS SUMMARY FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS DATA BASE MANAGEMENT

The following input-process-output (IPO) tables provide a general description of the data management processes associated with the standard analyses of NAEP data.

TABLE D-1. Creating and Loading Data Dictionaries

Inputs	Processes	Outputs
1. Assessment specifications	1. Structure and load exercise documentation and variable documentation files	1. Audited data dictionaries
2. Analysis specifications	2. Audit file contents	
3. Item documentation		
4. Background variable documentation		
5. Item classifications		
6. Reporting specifications		

TABLE D-2. Defining Assessment Respondent Data File Structure

Inputs	Processes	Outputs
1. Assessment content definitions (from data dictionaries)	1. Create file structure specifications	1. Respondent data file structure specifications
2. Assessment background factor definitions (from data dictionaries)		

TABLE D-3. Structuring and Loading Respondent Data Files

Inputs	Processes	Outputs
1. Respondent data file structure definitions (from Table D-2)	1. Structure file	1. Audited respondent data file
2. Respondent data file edit specifications (from data dictionaries)	2. Load file	2. Weight analysis reports
3. Field collected data (from scoring subcontractor)	3. Audit file	3. Sampling strata definition file
4. Sampling and supplementary data (from sampling and administration subcontractor)	4. Adjust weights	
5. Weight adjustment specifications	5. Derive sampling strata definitions	

TABLE D-4. Item-Level Data Analysis

Inputs	Processes	Outputs
1. Analysis specifications (from data dictionaries)	1. Calculate statistics for assessment items	1. Item-level performance data
2. Audited respondent data files (from Table D-3)	2. Load statistics to item performance data file	2. Analysis audit reports
3. Sampling strata definition file (from Table D-3)		3. Item performance data file

TABLE D-5. Item-Level Trend Analysis

Inputs	Processes	Outputs
1. Item-level performance data from previous cycle (from Table D-4)	1. Calculate change statistics	1. Item-level change data
2. Item-level performance data for current cycle (from Table D-4)	2. Load change statistics onto item performance data file	2. Change computation audit reports
		3. Item performance data file

TABLE D-6. Summary-Level Data Analysis/Reporting

Inputs	Processes	Outputs
1. Item-level performance data (from Table D-4)	1. Calculate item summary statistics	1. Summary performance data
2. Item classifications (from data dictionaries)	2. Calculate item summary change statistics	2. Summary performance change data
	3. Produce summary reports	3. Summary performance data reports

TABLE D-7. Item-Level Data Reporting

Inputs	Processes	Outputs
1. Item-level performance data (from Tables D-4 and D-5)	1. Merge item-level performance data, item documentation and background variable documentation	1. Item-level performance data reports
2. Reporting specifications (from data dictionaries)	2. Print reports according to reporting specifications	
3. Item documentation (from data dictionaries)		
4. Background variable documentation (from data dictionaries)		

APPENDIX E

AN OVERVIEW OF DATA MODELS, WITH EMPHASIS ON THE RELATIONAL MODEL

Data models are used to describe classes of information and the relationships among these classes. When working with complex sets of data, it is often useful to establish several levels of abstraction.

1. The level nearest the end user is called variously a view, a subscheme or an external schema, and represents the particular subset of data and relationships that one user (or a specific set of users) needs.

2. The level encompassing and integrating all the views across all user groups is called the conceptual schema, and represents the totality of the data and relationships.

3. The level dealing with the actual storage and manipulation of the data by the computer system is called the physical or internal schema.

Associated with views and conceptual schemas are data models specifying the types of relationships that the information, and the classes of information, have with each other.

There are three major data models currently in use in Data Base Management Systems. These are:

1. Hierarchical -- a tree structure, wherein the nodes of a tree represent entities (e.g., rooms), and the children of a node represent subordinate entities (e.g., furniture in a room).

2. Network -- a directed graph structure, wherein the nodes represent entities (e.g., types of automobiles, or types of tires), and arcs between nodes represent associations between entities (e.g., what tires fit on particular automobiles, or what automobiles will accommo-

date a particular tire).

3. Relational -- sets of ordered k-tuples[1] (in the set theoretic sense) used to denote relationships between entities (e.g., the ordered triple [EMPLOYEE, SALARY, MANAGER] describing salary and manager for employees, and concurrently both employees and managers for each salary, and employees and their salaries for each manager.

The relational model provides the descriptive power of both the hierarchical and network models, and does so with much simpler underlying concepts since the relational model is based on an underlying mathematical formalism, which provides a powerful theoretical basis for data manipulation operations. Five basic set manipulation operations provide a powerful language (relational algebra) for expressing queries against a relational data base. These are:

1. Union -- given two relations (both of arity k), the union of the relations is the set of k-tuples in either relation.

2. Set differences -- given two relations A and B (again both of arity k), the set difference A-B is the set of k-tuples in A but not in B.

3. Cartesian product -- given two relations A and B, having arity a and b, respectively, the Cartesian product A X B is the set of (a+b)-tuples whose first a components are an a-tuple in A, and whose last b components are a b-tuple in B.

4. Projection -- given a relation A of arity k, a projection of A is a new relation B, of arity less than k, whose components are tuples of A, with one or more specific component(s) removed. When relations are viewed as two-way tables, projection is the act of taking column subsets.

---

[1]The members of a relation are called tuples. K-tuples have k distinct pieces of information, or components. A relation consisting of k-tuples is said to have arity k.

5. Selection -- given a relation A of arity k, a selection on A is a new relation B, of arity k, whose tuples are all tuples of A. When relations are viewed as two-way tables, selection is the act of taking row subsets.

The relational data model offers several distinct advantages over the other data models.

1. Access to data does not require knowledge of navigation paths.

2. A system may expand easily; the insertion of a new data element is accomplished by determining the relationships the new data element has with existing data elements, and placing the new element in a new or existing relation, as appropriate.

3. Linkages among data are specified as combinations of the five basic set manipulation operations; no navigation logic is required.

4. The relationships between types of data are simple and easy to understand.

Both Principles of Database Systems (Ullman, 1980) and An Introduction to Database Systems (Date, 1977) provide an excellent discussion of data models, and the relational data model in particular.

APPENDIX F

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS RELEASE POLICY FOR CONFIDENTIAL DATA

The NAEP Assessment Policy Committee has established the following policy for release of confidential data (adopted June 1980):

> Data identifying individual schools, districts or states will not be provided by National Assessment to secondary users. The primary reasons for this prohibition are:
>
> 1. Historically, commitments pertaining to the confidentiality of information prohibit the release of data listing performance to any school, district or state.
>
> 2. A change in the current confidentiality commitments would result in the serious deterioration of the school "cooperation in the assessment" rate.
>
> Data will be provided to a secondary user desiring to link asessment data with other bases if:
>
> 1. The desired linkages are made by the National Assessment staff.
>
> 2. Aggregated data are provided for categories of schools, not individual schools or districts.
>
> Requests will be honored within schedule constraints. Data identifying individual schools, districts or states, or identifying groups of schools or districts by a listing of names will not be generated. Data pertaining to easily identifiable classifications of public and nonpublic education units will not be provided.

The linkage and generation of aggregate data will be funded by the requesting party. Audit of the aggregate data by a mutually acceptable person can be included in the request of the secondary user.

Any exception to this nonrelease of confidential data must be explicitly approved by the Assessment Policy Committee.

BIBLIOGRAPHY

Bachman, C.W. "Data Structure Diagrams," Data Base, vol. 1, no. 2, 1969.

Brooks, F.P. The Mythical Man-Month. Reading, Mass.: Addison-Wesley, 1975.

Chamberlin, D.D. "Relational Data-Base Management Systems," ACM Computing Surveys, vol. 8, no. 1, 1976.

Chamberlin, D.D. et al. "SEQUEL 2: A Unified Approach to Data Definition, Manipulation, and Control," IBM Journal of Research and Development, vol. 20, no. 6, 1976.

Cincom. OS TOTAL Reference Manual. Cincinnati, Ohio: Cincon Systems, 1978.

CODASYL. CODASYL Data Base Task Group April 71 Report. New York, N.Y.: ACM, 1971.

Cullinane. IDMS DML Programmer's Reference Guide. Wellesley, Mass.: Cullinane Corp., 1978.

Dahl, O.J. et al. Structured Programming. New York, N.Y.: Academic Press, 1972.

Date, C.J. An Introduction to Database Systems, 2nd ed. Reading, Mass.: Addison-Wesley, 1977.

Dijkstra, E.W. A Discipline of Programming. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1976.

Fry, J.P. and E.H. Sibley. "Evolution of Data-Base Management Systems," ACM Computing Surveys, vol. 8, no. 1, 1976.

Horowitz, E. and S. Sahni. Fundamentals of Data Structures. Potomac, Md.: Computer Science Press, 1976.

Jackson, M.A. Principles of Program Design. New York, N.Y.: Academic Press, 1975.

Kent, W. "Limitations of Record-Based Information Models," ACM Transactions on Database Systems, vol. 4, no. 1, 1979.

Kernighan, B. and P. Plauger. Software Tools. Reading, Mass.: Addison-Wesley, 1976.

Kernighan, B. and P. Plauger. The Elements of Programming Style. New York, N.Y.: McGraw-Hill, 1974.

Lorie, R.A. and J.F. Nilsson. "An Access Specification Language for a Relational Data Base System," IBM Journal of Research and Development, vol. 23, no. 3, 1979.

Major Informational Needs of National Assessment Audiences and Ways to Enhance the Assessment's Utility in Meeting Those Needs, no. 12-IP-51. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980.

Martin J. Computer Data-Base Organization. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1977.

Megowan, C.L. and J.R. Kelly. Top-Down Structured Programming Techniques. New York, N.Y.: Petrocelli/-Charter, 1975.

Michaels, A.S. et al. "A Comparison of the Relational and CODASYL Approaches to Data-Base Management," ACM Computing Surveys, vol. 8, no. 1, 1976.

MRI. System 2000 Reference Manual. Austin, Texas: MRI Systems Corp., 1978.

Myers, G.J. Composite/Structured Design. New York, N.Y.: Van Nostrand Reinhold Company, 1978.

Myers, G.J. Software Reliability. New York, N.Y.: John Wiley and Sons, 1976.

Olle, T.W. The Codasyl Approach to Data Base Management. New York, N.Y.: John Wiley and Sons, 1978.

Orr, K.T. Structured Systems Development. New York, N.Y.: Yourdon Press, 1977.

Shneiderman, B., ed. Database: Improving Usability and Responsiveness. New York, N.Y.: Academic Press, 1978.

Software AG. ADABAS Introduction. Reston, Va.: Software AG of North America, 1978.

Stonebraker, M.E. et al. "The Design and Implementation of INGRES," ACM Trans. on Database Systems, vol. 1, no. 3, 1976.

Tausworthe, R.C. Standardized Development of Computer Software. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1977.

Taylor, R.W. and R.L. Frank. "CODASYL Data-Base Management Systems," ACM Computing Surveys, vol. 8, no. 1, 1976.

Tsichritzis, D.C. and F.H. Lochovsky. Data Base Management Systems. New York, N.Y.: Academic Press, 1977.

Tsichritzis, D.C. and F.H. Lochovsky. "Hierarchical Data-Base Management: A Survey," ACM Computing Surveys, vol. 8, no. 1, 1976.

Ullman, Jeffrey D. Principles of Database Systems. Potomac, Md.: Computer Science Press, Inc., 1980.

Ward, Barbara. Major Informational Needs of National Assessment Audiences and Ways to Enhance the Assessment's Utility in Meeting Those Needs, no. 12-IP-51. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980.

Weinberg, G. The Psychology of Computer Programming. New York, N.Y.: Van Nostrand Reinhold Company, 1971.

Yourdon, E. How to Manage Structured Programming. New York, N.Y.: Yourdon, Inc., 1976.

Yourdon, E. and L. Constantine. Structured Design. New York, N.Y.: Yourdon, Inc., 1978.

Yourdon, E. Structured Walkthroughs. New York, N.Y.: Yourdon, Inc., 1978.

Yourdon, E. Techniques of Program Structure and Design. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1975.

Zloof, M.M. "Query-by-Example: A Data Base Language," IBM Systems Journal, vol. 16, no. 4, 1977.

Zook, W.K. et al. INGRES Reference Manual. Berkeley, Calif.: Department of EECS, University of California, 1977.